DTU Library

Joint overbooking and seat allocation for fare families

Hjorth, Robert; Fiig, Thomas; Bondoux, Nicolas; Larsen, Jesper

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Hjorth, R., Fiig, T., Bondoux, N., & Larsen, J. (2016). *Joint overbooking and seat allocation for fare families*.

Joint overbooking and seat allocation for fare families

Robert Hjorth[1], Thomas Fiig[2], Nicolas Bondoux[3], and Jesper Larsen[1]

[1] *Department of Management Engineering, Technical University of Denmark, Bygningtorvet, Building 424, DK-2800 Kongens Lyngby, Denmark*
[2] *Amadeus Airline IT, Lufthavnsboulevarden 14, DK-2770 Kastrup, Denmark*
[3] *Amadeus Airline IT, Amadeus S.A.S., 485 Route du Pin Montard, 06902 Sophia Antipolis Cedex, France*

July 15, 2016

Abstract

Revenue Management Systems (RMS) traditionally solve the seat allocation problem separately from the overbooking problem. Overbooking is managed by inflating the authorization levels obtained from seat allocation by various heuristics. This approach although suboptimal, is necessitated because of the complexity and dimensionality of the Dynamic Program (DP), which prohibits computation for realistic size problems.

We review several DP models developed for seat-allocation and overbooking over a time span of 40 years, reflecting changed business environments. In this report we link these models together by means of two transformations: The marginal revenue transformation of Fiig et al. [2010] and the equivalence charging scheme of Subramanian et al. [1999]. These transformations enable us to transform the joint seat allocation and overbooking problem for fare family fare structures into an equivalent independent demand model, which is readily solved. The resulting availability control can easily by implemented in existing RMS.

Keywords: Revenue Management Systems (RMS), overbooking, fare family fare structures

1 Introduction and Motivation

The introduction of Computer Reservation Systems (CRS) created the technology that enabled the ability to segment customers and hence created the opportunity for airlines to manage their seat inventory, which was the driver behind revenue management systems (RMS). The RMS were designed to support the airline business model by optimizing the price of the airline seats. This is still the case today although the RMS has undergone radical changes as the airline product has evolved over the past 40 years: point-to-point network (1980s); hub and spoke network (1990s); simplified fare structures (2000s); alliances, partnerships and mergers (2010s); to recent fare family fare structures.

Consequently, the RM research in the recent years has focused on forecasting and optimization for fare family structures, however no attention has been placed on overbooking. As it stands today, all commercial RM systems that we are aware of, split the optimization problem into two separate and consecutive steps: seat allocation and overbooking. This approach dates back to the birth of RM some 40 years ago. Although suboptimal, it was chosen due to the complexity and computational in-feasibility of solving the problems jointly. Thus, despite much progress in RM research - the state of the art for overbooking has basically not evolved.

The current practice is deficient is several ways. The overbooking methodology is inconsistent with the underlying fare structure. Overbooking models were built for fenced fare structures (typically one fare product is considered), while the airlines' fare structures have evolved into fare families. Further the current overbooking models ignore important factors that affect the overall overbooking level, such as: the demand level, the effects on class mix, the booking class specific refund costs, and booking class specific cancellation rates.

The purpose of the current paper is to solve the joint seat allocation and overbooking problem for fare families fare structures. Since this is largely an unexplored area we initially revisit the dynamic programming (DP) models applied in the context of seat allocation and overbooking with the purpose of extending and applying these models. We develop a general methodology based on linking the DP models, together by means of two transformations: The marginal revenue transformation (Fiig et al. [2010]) and the equivalence charging scheme (Subramanian et al. [1999]). These transformations enable us to transform the joint seat allocation and overbooking problem for fare family fare structures into an equivalent independent demand model, which may be solved exactly in a straightforward manner. We believe this methodology will be of great practical importance due to the prevalence of simplified fare structures.

In a later publication, we will share the results of our simulation studies were we compare of solution of the joint seat-allocation and overbooking problem with current industry practice and document significant revenue gain of 1% - 3% (Fiig et al. [2016]).

2 Literature review

The complete history on overbooking models since Rothstein and Stone [1967] will be too long to repeat here. We would like to refer the readers to very readable review papers by Ratliff [1998] and Rothstein [1985].

The evolution of revenue management over the last few decades as it relates to single leg seat allocation and overbooking is illustrated in Figure 3. The figure divides the existing literature according to two axes; Type of optimization model: seat-allocation only, overbooking only, or joint seat-allocation and overbooking; Type of demand model: independent demand (appropriate for fenced fare structures), and dependent demand (appropriate for fenceless- or fare-family fare structures). For completeness we include a brief literature review of the seat-allocation models, because some of these models and concepts will be applied in the current paper.

The optimization model are further categorized into static models and dynamic models. In the static models, we assume that demand for the multiple fare products book in a sequential order typically from low fare to high fare as we approach departure. In the dynamic models no assumptions of the booking order among the fare products are made.

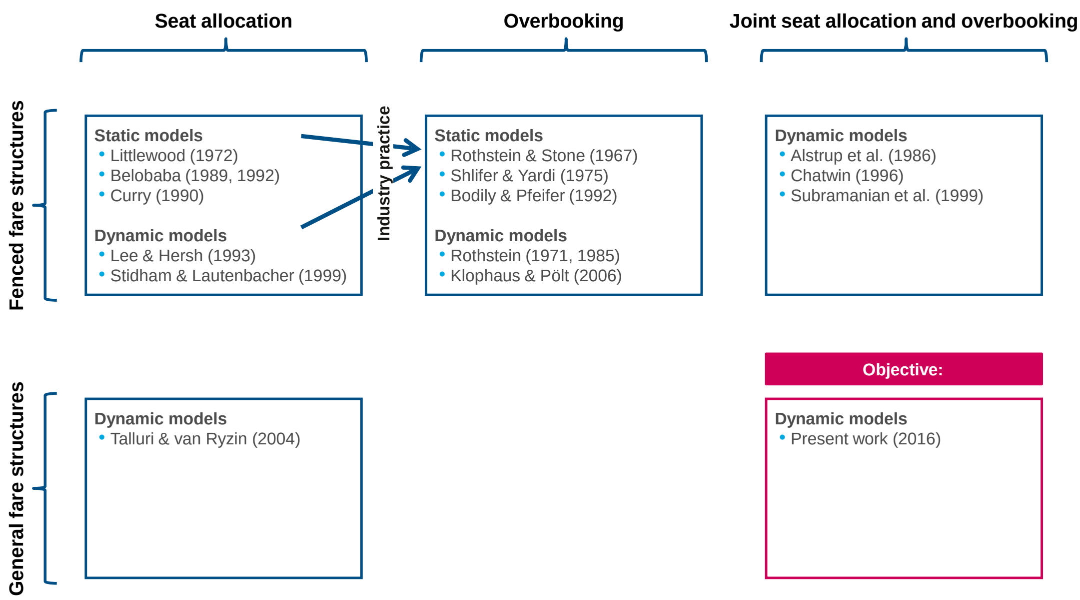

Figure 1: Literature overview

Seat-allocation: Independent demand

Before airline deregulation, routes, schedules, and fares were regulated in a way that ensured that the airlines earned a reasonable profit. Deregulation resulted in price volatility and a dramatic increase in the number of fares, as well as the introduction of fare structures with complex fare rules and restriction (among others min/max stay duration, advance purchase, Saturday/Sunday night restrictions, cancellation/refund flexibility) that differentiated the fare products.

The traditional revenue management systems (RMS), were built assuming a business environment of strong product differentiation, that enabled a critical simplification - to assume demand independence by fare product. This assumption was the foundation of RM systems built in the 1980s and 1990s that are still prevalent today even though the business environment has changed in response to competition from both incumbent airlines and low-cost carriers (LCC), see below.

The optimization of airline seat inventory to maximize revenue given multiple fare products on a single flight leg can be traced back to Littlewood [1972], who solved the fare class mix problem for two nested fare classes. Belobaba [1989] extended the nested seat allocation problem to multiple fare classes with the development of the Expected Marginal Seat Revenue (EMSR) heuristic. Curry [1990] later described optimal solutions for multiple nested fare classes.

Dynamic programming models were later applied by Lee and Hersh [1993] and Lautenbacher and Stidham [1999]. A fundamental result from these papers were that the optimal booking policy can be represented by a bid-price (BP), that depends only on a one-dimensional state variable (inventory, that is available seats), and time remaining to departure. Further the BP has the desirable properties: Inventory monotonicity (BP increases with less inventory), and time monotonicity (BP decreases with less time remaining to departure).

Seat-allocation: Dependent demand

In the 2000s, simplified fare structures started to emerge. The drivers came primarily from the competition from low cost carriers and the transparency in airline prices brought by the internet. The simplified fare structures invalidated the independent demand assumption, and caused the RM system to spiral down to the lowest booking class leading to tremendous loss of revenue, see Cooper et al. [2006].

Talluri and van Ryzin [2004] formulated a Dynamic Programming model for the case when consumer behavior is described by a general discrete choice model, and showed that only strategies on the efficient frontier are relevant in the optimization.

The development of the fare-adjustment theory Fiig et al. [2010], provided a theoretical basis for optimizing dependent demand, and has now become industry standard. The main result of the paper was to provide a marginal revenue transformation, that transforms an dependent demand model into an equivalent independent demand model (with adjusted fares and demands), that enable the continued use of the optimization methods and inventory controls developed for the traditional RMS.

Overbooking: General

As with the seat-allocation problem, we distinguish between static and dynamic models. In the static models, the time period is split into two stages a booking period, followed by a service period (following the terminology of Talluri and van Ryzin [2006]). In the booking period we accepted bookings up to given authorization levels, assuming demand is sufficient to exceed the authorization levels. In the service period the cancellations are realized considering its stochastic nature, and the surviving customers are serviced (assigned to the appropriate airline cabins, upgraded, down-graded) or denied boarded.

Typically only one fare class is considered in the static models, which represents the average fare used as empty seat cost.

The dynamic models, that we review later allows for more realism by considering booking events as well and in the case of joint seat-allocation and overbooking models also multiple fare products. The down-side is the in-tractability of these models.

The optimal overbooking level for both types of models is determined by balancing the risk and associated costs of empty seats and denied boardings respectively. Different optimization criteria, can be applied. The most common are given below:

i Maximize the economic profit (ticket revenue minus cost of denied boardings);

ii Max. allowed probability of a denied boarding;

iii Max. allowed proportion (service level) of denied boardings per number of flown passengers;

iv Min. total cost from spoilage + denied boardings.

Overbooking: Static models

Rothstein and Stone [1967] were among the first to model overbooking and apply their methodology in American Airlines (AA). They developed a static model to calculate the optimal booking level applying the max. probability criterion (ii) above. Their model was complicated by the fact that the industry practice at that time, allowed in addition to passengers with ordinary reservations; also teletype bookings, that arrived after reservations were closed and that had to be accepted due to interline agreements; as well as go-shows passengers, without a reservation, that arrive at departure with a valid ticket. Today these latter passenger types play a minor role and will not be considered in this paper.

Shlifer and Vardi [1975] considered three different extensions to the Rothstein and Stone model. The extensions regarded: 1) The single leg single class model, but applying different optimization criteria max. profit (i) and service level (iii). 2) The single leg but carrying two types of passengers with different cancellation rates. 3) Two leg network, three passenger types corresponding to the segments (two locals and a connection) with different cancellation rates. It is important to note that even in these simple generalizations (cases 2+3) compared to real-life airline networks with multiple fare classes, the optimal overbooking limits are quite complex accept/reject regions of the inventory state variables with limited practical applicability.

Bodily and Pfeifer [1992] extends the results of Shlifer and Vardi [1975] by relaxing the assumption of a constant survival rate to an assumption that allow time to departure dependent survival rates. The resulting decision rule remain unchanged applying the updated mean and standard deviation of the survival rates based on bookings in hand. Further they also considered an extension to allow for conditioning events (for example weather conditions) that cause cancellations of separate bookings to become correlated.

Overbooking: Dynamic models

Rothstein [1971] was the first to formulate a dynamic program for the overbooking problem. The model used is formulated as a non-homogenous Markov process. The objective was to determine the optimal booking policy that maximized the economic profit (criterion (i)). Although Rothsteins formulation was quite general and allows for various cancellation as well as distributions bookings eg. allowing group bookings or correlated cancellation events, the paper only considers Poisson distributed bookings and Binomial distributed cancellation. Under these assumptions the Rothstein model is equivalent to the model from Subramanian et al. [1999], which is much simpler to analyze, see literature review below.

Klophaus and Pölt [2006] extended the static Rothstein and Stone [1967] model by applying a dynamic spoilage cost that better represents the different passengers willingness to pay (WTP) across the booking horizon. They propose to use the lowest available fare level as a proxy for the passengers WTP. Simulation studies on real Lufthansa data, testing all four optimization criteria (i)-(iv) displayed minor incremental revenue gains over the static model.

Joint seat allocation and overbooking models Dynamic models

Alstrup and Boas [1984] formulated the joint allocation and overbooking problem for two passenger types (business and leisure) as a Markovian non-homogenous sequential decision process, that were solved using two-dimensional dynamic program for the optimal overbooking policy. Their model can be considered as a generalization of the Rothstein [1971] to solve the problem with two passenger types. The objective was to determine the optimal booking policy that maximize the economic profit (criterion (i)).

Although the Boas and Alstrup formulation was quite general and allows for various booking as well as cancellation distributions, the paper only considers Poisson distributed bookings and Binomial distributed cancellation. The paper considers the complexity introduced by actual aircraft configurations involving position of the cabin divider, up-gradings, downgradings. Subramanian et al. [1999], later considered a generalization of the Boas and Alstrup model to allow for multiple fare classes. However ignoring the practical configuration complexities considered by Alstrup and Boas [1984], and the extensions to include refund by Subramanian et al. [1999], the two models are equivalent.

Chatwin [1996] formulated the joint seat allocation and overbooking problem using a continuous space, discrete time, dynamic programming model for multiple fare classes. Given the assumptions that during each time-slice, bookings can only occur in one class, and the number of cancellations are proportional to the number of reservations; he proved that a booking limit policy is optimal. The booking limits for each fare class depends on the number of reservations made in the other classes. For two fare classes specifically Chatwin [1996] proves the booking limit monotonicity property.

It should be noted that in the limit where time slices become so thin, that at most one booking event/cancellation event can occur, Poisson distributed demand and Binomial distributed cancellations, satisfy the assumptions above.

Subramanian et al. [1999] consider the joint seat-allocation and overbooking (allowing for no-show and refunds) problem for a single leg, single cabin with multiple fare classes. They extend the results of Lee and Hersh [1993] for one class problem (or equivalently multiple fare classes, where all classes having identical cancellation probabilities). Analogous to Chatwin they prove that a booking limit policy is optimal and that the bidprice satisfy the monotonicity conditions: inventory monotonicity and time monotonicity, see above.

3 Overview of models

The aim of this paper is to solve the seat allocation and overbooking problems jointly for simplified fare structures. For this purpose, we shall first present an overview of selected existing dynamic programming (DP) models for seat allocation and overbooking. Following this review, we develop the general methodology for solving the seat allocation and overbooking problems jointly for fare family fare structures.

In this section, we show how the models of Lee and Hersh [1993], Subramanian et al. [1999] and Talluri and van Ryzin [2004] link neatly together, and in fact can be derived from each other through application of the transformations of Subramanian et al. [1999] and Fiig et al. [2010]. The relationship between the selected models and transformations, as well as to the present work, is depicted in Figure 2. Similar to Figure 3 in the literature review, the models in Figure 2 are grouped according to two axes: type of optimization model and type of demand model. A number of basic properties are listed for each model: dimensionality of the state variable, monotonicity properties of the bid-price and acceptance criteria. An arrow between two models represents exact equivalence of the models through application of the transformation shown on top of the arrow.

For consistency, the models are re-named according to type of demand model (D and I for dependent and independent demand models respectively) and type of optimization model (OB for joint seat allocation and overbooking models). In addition, joint seat allocation and overbooking models are named according to how the cancellation cost is treated; as a total cost (TC) of all simultaneous cancellations (paid at the time of refund) or as an expected unit cost (UC) of cancellation for one incremental passenger (treated in the optimization at the time of booking). The transformations are abbreviated to ECT (Equivalence Charging Transformation) and MRT (Marginal Revenue Transformation).

In the following, we will describe and formulate the models chronologically and show how they link together. Independent demand models will be considered first. We show how to extend Model I to include cancellations and overbooking, and then prove equivalence of Model I+OB (TC) and Model I+OB (UC) by applying the Equivalence Charging Transformation. The ECT transforms cancellation cost from being treated at the time of refund to the time of booking by subtracting from the fare the expected unit cost of cancellation inflicted by the passenger

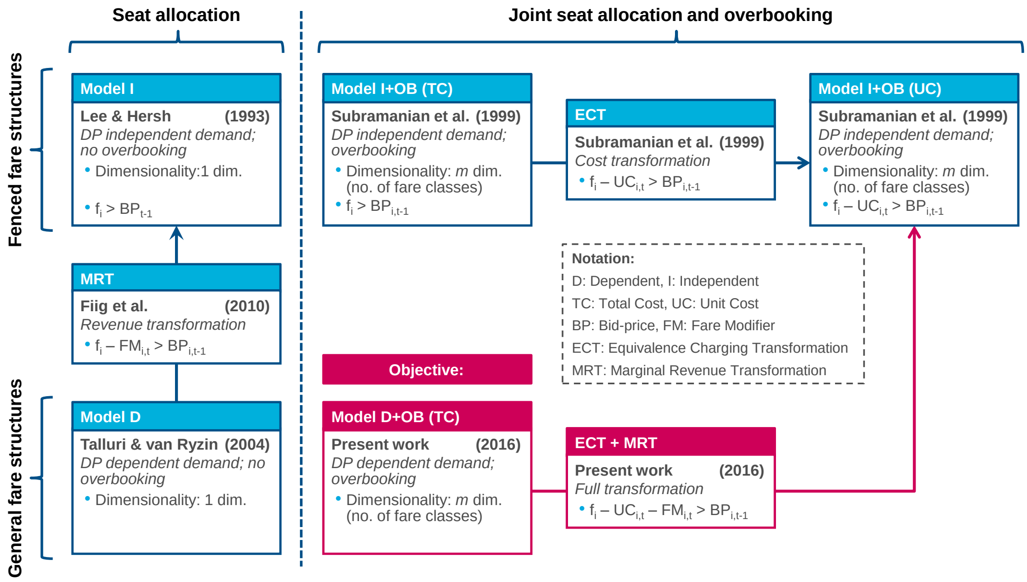

Figure 2: Overview of DP models

requesting to book the fare. As we shall see, the advantage of treating cancellation costs up front is that the dimensionality of the DP can be reduced.

Subsequently, we consider the dependent demand models and show equivalence of Model D and Model I by applying the Marginal Revenue Transformation. The MRT transforms a dependent demand model into an equivalent independent demand model by calculating a fare modifier to be subtracted from the original fare. This transformation enables the continued use of traditional RMS for fare structures with dependent demand by making it possible to use Model I even on fare structures assuming dependent demand.

Finally, we develop a methodology to transform Model D with any type of demand and fare structure into Model I+OB (UC) by extending the ECT to dependent demand models, realizing that the MRT also applies to cost and applying the two extended transformations in sequence. We thereby show equivalence of Model D+OB (TC) and Model I+OB (UC).

Throughout the rest of the paper, notation will be an adapted form of the notation used in Subramanian et al. [1999].

Notation for fenced and fenceless fare structures

Assume we have m fare classes and a state variable $\mathbf{x} = (x_1, \ldots, x_m)$ in m dimensions to keep track of the number of bookings in each fare class. The pure seat allocation models also assume m fare classes, but only need a one-dimensional state variable $x = \sum_{i=1}^{m} x_i$ to keep track of the total number of bookings. Fares are denoted by f_i for fare class i, and are listed in decreasing fare order: $f_i > f_{i+1}$, $i = 1, \ldots, m-1$.

For the sake of simplicity, we develop the dynamic programming models for a single leg, single cabin flight with capacity C. The booking horizon is sliced into $T+1$ stages, numbered in reverse chronological order, $t = T, T-1, \ldots, 1, 0$ with departure occurring at $t = 0$. In each stage, we make the assumption of exactly one event occurring (booking request, cancellation or no event), and ensure it is satisfied by choosing T to be sufficiently large.

A policy p is a set of open fare classes. We denote demand for fare class i given that policy p is offered in stage t by $d_{i,p,t}$, where $d_{i,p,t} = 0$ if $i \notin p$ or p is the null set where no fare classes are offered. The probability of a request in policy p in stage t is given by $\sum_{i\in p} d_{i,p,t}$ and the probability of no request by $1 - \sum_{i\in p} d_{i,p,t}$. Aggregated demand for policy p is denoted by $D_{p,t} = \sum_{i\in p} d_{i,p,t}$. When policies and fare classes are identical, we remove the policy index and let $d_{i,t}$ denote demand for fare class i in stage t and assume it is independent of the current number of bookings in any fare class. We denote the cancellation rate by $q_{i,t}$ and associate with it a refund of c_i for fare class i in stage t.

We will present model-specific assumptions as they are introduced.

Review of Model I: Lee and Hersh [1993]

Assumptions: Seat allocation DP with 1-dim state variable, demand independence by fare product, and no cancellations

Lee and Hersh considered a traditional fenced fare structure and independent product-oriented demand without cancellations in Lee and Hersh [1993]. They considered a single flight leg and a single cabin, and developed a DP model to solve the seat allocation problem.

The DP model is solved for all stage-state combinations, with the objective to maximize expected remaining revenue of operating the system over the booking horizon from $t = T$ to $t = 0$. Let $TR_t(x)$ denote the expected remaining revenue in state x from stage t to 0. Note that the state variable is one-dimensional although there are m fare classes (as long as cancellations are not considered). The boundary conditions are $TR_t(C) = 0$ (no further obtainable revenue when capacity is reached) and $TR_0(x) = 0$ (no further revenue or cost at time of departure).

An accept/reject decision in stage t has to be made whenever a booking request arrives. The value of the request is the fare f_i, and we accept it if $f_i + TR_{t-1}(x+1) > TR_{t-1}(x)$, i.e. if the expected remaining revenue associated with accepting the booking is larger than the expected remaining revenue of rejecting it. The first term in the model represents the sum over all independent revenue streams and the associated accept/reject decisions. If we accept a request, we continue with $x+1$ bookings in the next stage, while a reject decision leaves the number of bookings unaltered. The second term in the model represents the expected remaining revenue when there is no booking request, in which case x is also left unaltered. The Bellman equation can be formulated as

$$TR_t(x) = \sum_{i=1}^{m} d_{i,t} \cdot \max\{f_i + TR_{t-1}(x+1), TR_{t-1}(x)\} + \left(1 - \sum_{i=1}^{m} d_{i,t}\right) TR_{t-1}(x)$$

where we can simplify somewhat by subtracting $\sum_{i=1}^{m} d_{i,t} \cdot TR_{t-1}(x+1)$ inside the maximization and adding it back outside. This yields Model I

$$TR_t(x) = \sum_{i=1}^{m} d_{i,t} \cdot \max\{f_i - BP_{t-1}(x), 0\} + TR_{t-1}(x) \tag{1}$$

where $BP_{t-1}(x) = TR_{t-1}(x) - TR_{t-1}(x+1)$ denotes the bid-price. We accept a booking if $f_i - BP_{t-1}(x) > 0$ and reject it otherwise.

Lee & Hersh proved monotonicity of the bid-price in inventory and time in Lee and Hersh [1993]. The bid-price is non-decreasing in x (non-increasing in remaining inventory) for constant t, i.e. $BP_t(x) \leq BP_t(x+1)$ (bid-price increases with more bookings / decreases with more remaining capacity), and non-decreasing in t for constant x, i.e. $BP_{t-1}(x) \leq BP_t(x)$ (bid-price decreases towards departure). The property of inventory monotonicity guarantees the existence of an optimal booking limit for each fare class i in each stage t, since there will be a maximum x for which the acceptance criteria $f_i - BP_{t-1}(x) > 0$ is satisfied (if $f_i - BP_{t-1}(0) \leq 0$ then the optimal booking limit is to close fare class i completely in stage t). Likewise, the property of time monotonicity guarantees the existence of an optimal time for when to close each fare class i for a given state x.

Review of Model I+OB (TC): Subramanian et al. [1999]

Assumptions: Joint seat allocation and overbooking DP with m-dimensional state variable, demand independence by fare product, cancellations and overbooking, and refund considered at time of cancellation

Subramanian et al. [1999] extended Model I to include cancellations, no-shows, refunds and overbooking (in this paper, we exclude no-shows for the sake of notational simplicity, but without loss of generality). The basic assumptions are the same as in Lee and Hersh [1993] – they assume a traditional fenced fare structure and independent product-oriented demand on a single flight leg served by a single-cabin aircraft.

In order to have class-dependent cancellation rates and refunds, the DP model is extended to m dimensions to keep track of the number of bookings in each fare class through the m-dimensional state vector $\mathbf{x} = (x_1, \ldots, x_m)$. In the general case of the model, demand and cancellation rates are allowed to depend on the full state vector $\mathbf{x}$ through the functions $d_{i,t}(\mathbf{x})$ and $q_{i,t}(\mathbf{x})$. However, this generality adds little additional value in practical applications, so we derive the model with less general assumptions. Demand $d_{i,t}$ is assumed to depend on fare class and time, but not on bookings in hand. Similarly, cancellation rates $q_{i,t}$ are assumed to depend on fare class and time. In addition, the number of cancellations in each fare class is assumed to be proportional to the number of bookings in hand, i.e. we expect $q_{i,t}x_i$ cancellations in fare class i in stage t (as a consequence each passenger cancels independently of each other).

The acceptance criteria for a new booking request in fare class i is $f_i - BP_{i,t-1}(\mathbf{x}) > 0$ in stage t, where $BP_{i,t-1}(\mathbf{x}) = TR_{t-1}(\mathbf{x}) - TR_{t-1}(\mathbf{x}+\mathbf{e}_i)$ is now the i'th component of the bid-price in m dimensions ($\mathbf{e}_i$ is the unit vector of all zeroes and a one at index i). Model I+OB (TC) is structured in the same way as Model I, with the first term representing the sum of accept/reject decisions for all independent revenue streams in the event of a booking request and the last term representing the expected remaining revenue when no event occurs (booking or cancellation). In addition, we now need a term to handle the expected cancellations in each fare class. In the event of a cancellation in fare class i, which has the probability $q_{i,t}x_i$ of occurring, we continue with $\mathbf{x} - \mathbf{e}_i$ bookings and refund the amount c_i. The Bellman equation for Model I+OB (TC), including cancellations, refunds and overbooking, is

$$\begin{aligned} TR_t(\mathbf{x}) = \quad & \sum_{i=1}^{m} d_{i,t} \cdot \max\left\{f_i - BP_{i,t-1}(\mathbf{x}), 0\right\} + \sum_{i=1}^{m} q_{i,t}x_i \left(-c_i + TR_{t-1}(\mathbf{x} - \mathbf{e}_i)\right) + \\ & \left(1 - \sum_{i=1}^{m} q_{i,t}x_i\right) TR_{t-1}(\mathbf{x}) \end{aligned} \tag{2}$$

Since overbooking is allowed, denied boardings also have to be considered. Denote the overbooking penalty function by $\pi(\cdot)$ where π is convex and non-decreasing function describing the denied boarding cost. The boundary condition for Model I+OB (TC) is $TR_0(\mathbf{x}) = -\pi\,(x)$.

In the general case of Model I+OB (TC) with the m fare classes having different cancellation rates, the bid-price satisfies inventory monotonicity but not time monotonicity.

Model I+OB (TC) solves the seat allocation and overbooking problems jointly, and is an important improvement over Model I since it includes cancellations, no-shows, refunds and overbooking. However, the Bellman equation is m-dimensional in order to handle class-dependent cancellation rates, which for all practical purposes is infeasible to solve in real time for realistic airline problems with a large m. Model I+OB (TC) can be reduced to one dimension provided the cancellation rates and refunds being independent of fare class. In practice this is unlikely to be true.

Review of Equivalence Charging Transformation (ECT): Subramanian et al. [1999]

Transformation of cancellation cost

To overcome the computational complexity of having m dimensions, Subramanian et al. [1999] developed a clever method to reformulate the problem, that allowed a reduction to a one-dimensional state variable with less restrictive assumptions. In Model I+OB (TC), the refund cost is treated when the cancellation event occurs. The Equivalence Charging Transformation (ECT) transforms the Bellman equation into an equivalent equation, in which the expected cancellation cost is instead assessed at time of booking. The total expected cost of cancellation in hand from t to departure is denoted by $TC_t(\mathbf{x})$ and defined as a recursion formula. If a cancellation in fare class i occurs in stage t, the amount c_i is refunded and the number of bookings in hand will be $\mathbf{x} - \mathbf{e}_i$ going in to stage $t-1$. If no cancellation occurs in stage t, the expected remaining cancellation cost is given by $TC_{t-1}(\mathbf{x})$. We will assume (as also done in Subramanian et al. [1999]) that

1. Cancellations in one fare class being independent of number of bookings in other fare classes.

2. Cancellations are independent of each other.

The recursion formula becomes:

$$TC_t(\mathbf{x}) = \sum_{i=1}^{m} q_{i,t} x_i (c_i + TC_{t-1}(\mathbf{x} - \mathbf{e}_i)) + \left(1 - \sum_{i=1}^{m} q_{i,t} x_i\right) TC_{t-1}(\mathbf{x}) \tag{3}$$

where $TC_0(\mathbf{x}) = 0$. Note that $TC_t(\mathbf{x})$ is positive by convention. By adding it back to $TR_t(\mathbf{x})$, from which cancellation cost is subtracted in the form of c_i, we can calculate the gross revenue (i.e. revenue before subtracting cancellation costs) of operating the system from t to departure by means of the equation $TB_t(\mathbf{x}) = TR_t(\mathbf{x}) + TC_t(\mathbf{x})$. By inserting $TB_t(\mathbf{x}) - TC_t(\mathbf{x})$ in place of $TR_t(\mathbf{x})$ in (2) we obtain:

$$\begin{aligned} TB_t(\mathbf{x}) = & \sum_{i=1}^{m} d_{i,t} \cdot \max\{f_i - [TC_{t-1}(\mathbf{x} + \mathbf{e}_i) - TC_{t-1}(\mathbf{x})] - BP_{i,t-1}(\mathbf{x}), 0\} \\ & + \sum_{i=1}^{m} q_{i,t} x_i \cdot TB_{t-1}(\mathbf{x} - \mathbf{e}_i) + \left(1 - \sum_{i=1}^{m} q_{i,t} x_i\right) TB_{t-1}(\mathbf{x}) \end{aligned} \tag{4}$$

where $BP_{i,t-1}(\mathbf{x}) = TB_{t-1}(\mathbf{x}) - TB_{t-1}(\mathbf{x} + \mathbf{e}_i)$.

Applying the assumptions above allows us to calculate (3) as a sum of independent expected remaining cancellation costs for each fare class; $TC_t(\mathbf{x}) = \sum_{i=1}^{m} TC_{i,t}(x_i)$, where $TC_{i,t}(x_i)$ is given as the i'th component of the recursion above

$$TC_{i,t}(x_i) = q_{i,t} x_i (c_i + TC_{i,t-1}(x_i - 1)) + (1 - q_{i,t} x_i) \cdot TC_{i,t-1}(x_i)$$

where $TC_{i,0}(x_i) = 0$ for all i. From this one-dimensional recursion formula we can define the expected unit cost of cancellation attributable to one incremental booking in fare class i as $UC_{i,t}(x_i) = TC_{i,t-1}(x_i + 1) - TC_{i,t-1}(x_i)$. It follows that each of the functions $UC_{i,t}(x_i)$ satisfy the recursion

$$UC_{i,t}(x_i) = q_{i,t-1} c_i + (1 - q_{i,t-1}) \cdot UC_{i,t-1}(x_i) + q_{i,t-1} x_i \cdot (UC_{i,t-1}(x_i - 1) - UC_{i,t-1}(x_i))$$

for $t \geq 2$ and with $UC_{i,1}(x_i) = 0$. Note that the time indices are moved by one in the definition of $UC_{i,t}(x_i)$ since a cancellation happens no earlier than the stage immediate after the booking is made. We now apply assumption (ii) of independence of cancellations to see that $UC_{i,t-1}(x_i - 1) = UC_{i,t-1}(x_i)$, i.e. that the expected unit cost of cancellation from one incremental booking in fare class i in stage t is identical for x_i and $x_i - 1$ bookings in hand. Not surprisingly, the consequence of assuming independent cancellations is that we can eliminate x_i from the recursion. Hence, the expected unit cost of cancellation becomes:

$$UC_{i,t} = q_{i,t-1} c_i + (1 - q_{i,t-1}) \cdot UC_{i,t-1} \tag{5}$$

for $t \geq 2$ and with $UC_{i,1} = 0$. Finally we replace each component of the vector $TC_{t-1}(\mathbf{x} + \mathbf{e}_i) - TC_{t-1}(\mathbf{x})$ in (4) with the unit cost $UC_{i,t}$ to obtain Model I+OB (UC), see the following paragraph.

The ECT in itself does no more than transform the Bellman equation of Model I+OB (TC) into the equivalent Model I+OB (UC), see below. below). However, the impact of the ECT is significant, since it will allow reduction of the Bellman equation to one dimension with less restrictive assumptions.

Review of Model I+OB (UC): Subramanian et al. [1999]

Assumptions: Joint seat allocation and overbooking DP with m-dimensional state variable. Fenced fare structures. Demand independence by fare product, cancellations, overbooking, and refund cost considered at time of booking

Models I+OB (TC) and I+OB (UC) are equivalent. The only difference being how cancellation costs are treated - at the time of booking vs. at the time of cancellation. Treating cancellation costs at the time of booking result in the value of the booking being assessed as the fare f_i minus the expected unit cost $UC_{i,t}$ of cancellation. When making the accept/reject decision, this value is compared to the bid-price $BP_{i,t-1}(\mathbf{x})$, and the acceptance criteria becomes $f_i - UC_{i,t} - BP_{i,t-1}(\mathbf{x}) > 0$. Treating cancellation cost up front also means that the refund c_i is no longer included in the Bellman equation. The Bellman equation for Model I+OB (UC) becomes:

$$
\begin{aligned}
TB_t(\mathbf{x}) = \quad & \sum_{i=1}^{m} d_{i,t} \cdot \max\{f_i - UC_{i,t} - BP_{i,t-1}(\mathbf{x}), 0\} + \sum_{i=1}^{m} q_{i,t} x_i \cdot TB_{t-1}(\mathbf{x} - \mathbf{e}_i) + \\
& \left(1 - \sum_{i=1}^{m} q_{i,t} x_i\right) TB_{t-1}(\mathbf{x})
\end{aligned}
\tag{6}
$$

where $BP_{i,t-1}(\mathbf{x}) = TB_{t-1}(\mathbf{x}) - TB_{t-1}(\mathbf{x} + \mathbf{e}_i)$. Denied boardings are considered through the boundary condition $TB_0(\mathbf{x}) = -\pi(x)$, as before.

The ECT preserves monotonicity properties, so the bid-price in Model I+OB (UC) still satisfies inventory monotonicity but not time monotonicity.

As we have seen, allowing cancellation rates and refunds to be class-dependent results in the Bellman equation being m-dimensional, which is infeasible to solve for real sized problems. With realistic assumptions of passengers cancelling independently of each other, the ECT made it possible to evaluate the expected net revenue of accepting a booking request up front instead of waiting for the passenger to cancel to see the refund cost in the model. By further assuming $q_{i,t} = q_t \ \forall i$, Subramanian et al. [1999] were able to reduce the DP to one dimension.

Review of Model D: Talluri and van Ryzin [2004]

Assumptions: Seat allocation DP with one-dimensional state variable. General fare structure. Demand dependence by fare product, and no cancellations.

Talluri & van Ryzin formulated a DP model for a general discrete choice model in Talluri and van Ryzin [2004], in which demand depends on the policy offered expressed as a general choice model. The objective of the DP model is again to maximize expected remaining revenue of operating the system from t to departure, but the control problem is different from the independent demand problem. The general DP model has to maximize the expected remaining revenue $TR_t(x)$ over all policies. The probability of a request in policy p in stage t is given by $\sum_{i \in p} d_{i,p,t}$ and the probability of no request by $1 - \sum_{i \in p} d_{i,p,t}$. Any request for an open fare class is accepted, and will generate the revenue of $f_i + TR_{t-1}(x+1)$, since we will have one more booking when entering stage $t-1$. In case of no requests for open fare classes, x is left unaltered. The Bellman equation for Model D can therefore be stated as

$$
\begin{aligned}
TR_t(x) &= \max_p \left\{ \sum_{i \in p} d_{i,p,t}\left(f_i + TR_{t-1}(x+1)\right) + \left(1 - \sum_{i \in p} d_{i,p,t}\right) TR_{t-1}(x) \right\} \\
&= \max_p \left\{ \sum_{i \in p} d_{i,p,t}\left(f_i - BP_{t-1}(x)\right) \right\} + TR_{t-1}(x)
\end{aligned}
\tag{7}
$$

where we again find the bid-price $BP_{t-1}(x) = TR_{t-1}(x) - TR_{t-1}(x+1)$. Similar to Model I, the boundary conditions are $TR_t(C) = 0$ and $TR_0(x) = 0$ since cancellations and overbooking are not considered.

Talluri and van Ryzin prove identical monotonicity properties for Model D as Lee and Hersh proved for Model I. The bid-price is non-decreasing in x for constant t, i.e. $BP_t(x) \leq BP_t(x+1)$, and non-decreasing in t for constant x, i.e. $BP_{t-1}(x) \leq BP_t(x)$. These properties guarantee existence of an optimal solution to the problem, which takes form as a policy of which fare classes to open at a certain stage t given x bookings in hand.

Review of Marginal Revenue Transformation (MRT): Fiig et al. [2010]

Transformation of dependent demand model into an equivalent independent demand model.

Fiig et al. [2010] demonstrate how to transform a dependent demand model into an equivalent independent demand model. Since we shall need the MRT later, we will review the method here. The significance of the MRT is two-fold. The theoretical significance is that it shows solution equivalence of Model D and Model I. And the practical application is that it provides a basis for optimizing dependent demand in models assuming independent demand, thus enabling the continued use of the optimization methods developed for traditional airline RMS.

In the dependent model, optimization is performed over all policies. Revenue and demand depend on the policy being offered. In stage t, demand for policy p is $D_{p,t} = \sum_{i\in p} d_{i,p,t}$ and revenue is $R_{p,t} = \sum_{i\in p} d_{i,p,t} f_i$. In order to find the optimal policy, we can plot $D_{p,t}$, $R_{p,t}$ in a scatterplot for all policies. This will trace out an efficient frontier. We only have to consider policies on the efficient frontier (all points falling below the efficient frontier provides less revenue for the same capacity consumption and are clearly inefficient).

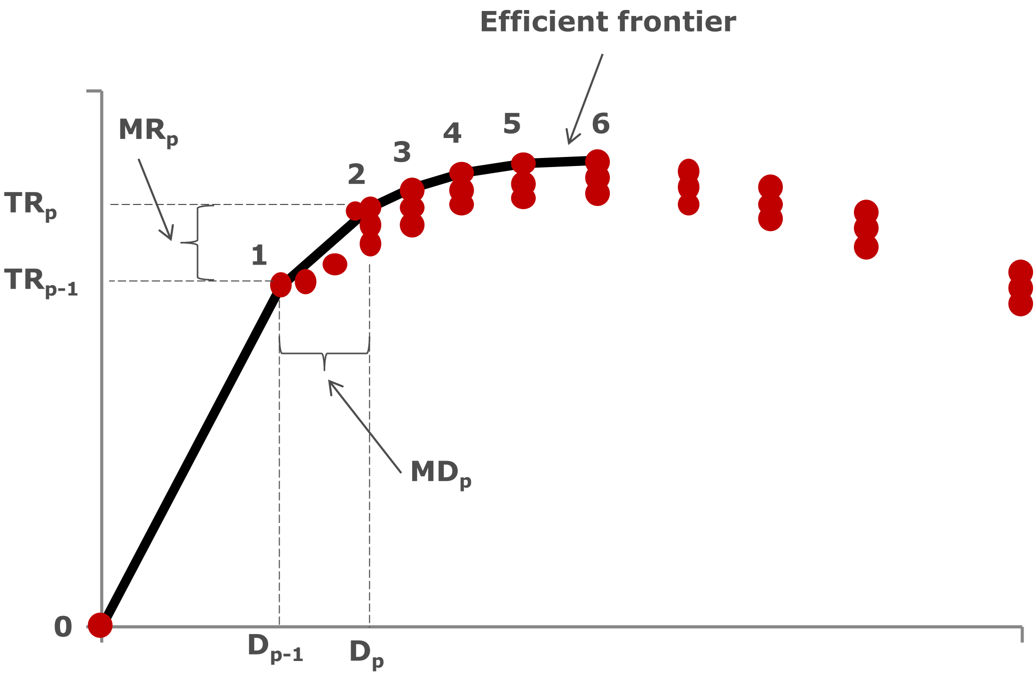

Figure 3: Marginal Revenue Transformation

The brilliant idea of Fiig et al. [2010] was to create an equivalent efficient frontier consisting of transformed fare classes with independent demand. This is achieved by a two-step approach. The first step is to create a virtual policy with independent demand for each policy on the efficient frontier. Independent demand for the virtual policies is set equal to the marginal demand $MD_{p,t}$ obtained by opening the next policy with dependent demand on the efficient frontier. The adjusted fare for the virtual policy is likewise set equal to the marginal revenue $MR_{p,t}$. The marginal demand and marginal revenue of opening policy p instead of $p-1$ is

$$MD_{p,t} = D_{p,t} - D_{p-1,t}$$

$$MR_{p,t} = (R_{p,t} - R_{p-1,t})/(D_{p,t} - D_{p-1,t}). \tag{8}$$

where $p = 1, \ldots, k$ denotes the k policies along the efficient frontier. By definition, the policies trace out an efficient frontier equivalent to the one traced out by the dependent demand policies.

The second step of the MRT is to map the policies back to the original m fare classes. As noted in Fiig et al. [2010], this is only possible if the policies along the efficient frontier are nested. In that case, the policies can be mapped to fare classes by assigning to each fare class opening for the first time $MR_{p,t}$ as an adjusted fare $MR_{i,t} = f_i - FM_{i,t}$ and $MD_{p,t}$ as demand. Here we for convenience have introduced the fare modifier $FM_{i,t}$ as the fare adjustment. By definition, this procedure generates an equivalent fare structure, now constituted by transformed fare classes with independent demand.

The transformation will allow us to solve the optimization problem defined by Model D using Model I by applying transformed demand and fares.

4 Present work: Joint overbooking and seat allocation for fare families

Model D+OB (TC): Joint overbooking and seat allocation for fare families

Assumptions: Joint seat allocation and overbooking DP with m-dimensional state variable. Fare family fare structure. Demand dependence by fare product, cancellations, overbooking, and refund at time of cancellation.

We derive the DP model for the general case with fare family fare structure, and therefore need to formulate the maximization over policies. The demand for fare class i given that policy p is offered in stage t is again denoted by $d_{i,p,t}$ and the probability of a request in policy p in stage t can be computed by $\sum_{i \in p} d_{i,p,t}$. The number of cancellations in each fare class is assumed to be proportional to the number of bookings in hand, i.e. we expect $q_{i,t} x_i$ cancellations in fare class i in stage t.

The first term of the Bellman equation maximizes revenue over the policies, and is similar to the maximization term of Model D. The second term considers a cancellation event in fare class i, in which case the amount c_i is refunded and we continue with $\mathbf{x} - \mathbf{e}_i$ bookings. Finally, the third term considers the event of no booking or cancellation. The Bellman equation of Model D+OB (TC) is

$$\begin{aligned} TR_t(\mathbf{x}) = \quad & \max_p \left\{ \sum_{i \in p} d_{i,p,t} \left(f_i - BP_{i,t-1}(\mathbf{x}) \right) \right\} + \sum_{i=1}^{m} q_{i,t} x_i \left(-c_i + TR_{t-1}(\mathbf{x} - \mathbf{e}_i) \right) + \\ & \left(1 - \sum_{i=1}^{m} q_{i,t} x_i \right) TR_{t-1}(\mathbf{x}) \end{aligned} \tag{9}$$

where $BP_{i,t-1}(\mathbf{x}) = TR_{t-1}(\mathbf{x}) - TR_{t-1}(\mathbf{x} + \mathbf{e}_i)$ and the boundary condition $TR_0(\mathbf{x}) = -\pi(x)$ is the overbooking penalty function.

Similar to the other m-dimensional models derived in this section, the bid-price in Model D+OB (TC) exhibits inventory monotonicity but not time monotonicity.

Model D+OB (UC): Joint overbooking and seat allocation for fare families

Assumptions: Joint seat allocation and overbooking DP with n-dimensional state variable. Fare family fare structure. Demand dependence by fare product, cancellations, overbooking, and refund at time of booking.

We use the same notation as above, but add an index j to denote the n fare families. We let $f_{j,p} = \min_{i \in j, i \in p} \{f_i\}$ denote the lowest offered fare from family j in policy p and $UC_{j,p,t}$ denote the expected unit cost of cancellation resulting from one additional booking in the lowest offered fare class from family j in policy p at stage t. Demand for fare class i in stage t given that policy p is offered is denoted by $d_{i,p,t}$, which is equal to zero if $i \notin p$. Aggregated demand for fare family j in stage t given that policy p is offered can be computed by $D_{j,p,t} = \sum_{i \in j} d_{i,p,t}$.

For the general model, we assume cancellation rates $q_{j,t}$ to depend on which fare family j the booking is made in, and we therefore need the state vector to be n-dimensional to keep track of the number of bookings in each fare family.

First, we formulate the general fare family model for n fare families directly in the unit cost based version, since performing the ECT on a fare family version of Model D+OB (TC) is straightforward. The Bellman equation for the general fare family model (Case nE) is

$$\begin{aligned} TB_t(\mathbf{x}) = \max_p & \left\{ \sum_{j=1}^{m} D_{j,p,t} \left(f_{j,p} - UC_{j,p,t} - BP_{j,t-1}(\mathbf{x}) \right) \right\} \\ & + \sum_{j=1}^{m} q_{j,t} x_j \cdot TB_{t-1}(\mathbf{x} - \mathbf{e}_j) + \left(1 - \sum_{j=1}^{m} q_{j,t} x_j \right) TB_{t-1}(\mathbf{x}) \end{aligned} \tag{10}$$

where $BP_{j,t-1}(\mathbf{x}) = TB_{t-1}(\mathbf{x}) - TB_{t-1}(\mathbf{x} + \mathbf{e}_j)$. Note how we have leveraged the fundamental property of fare families being unrestricted to formulate the general fare family model in the exact same form as (**??**), by letting each policy consist of one (or none) fare class from each fare family.

1 Dim. Approximation. Model D+OB (UC): Joint overbooking and seat allocation for fare families

Assumptions: Joint seat allocation and overbooking DP with n-dimensional state variable. Fare family fare structure. Demand dependence by fare product, cancellations, overbooking, and refund at time of booking.

Finally we reduce dimensionality to one by assuming the cancellation rate q_t to be identical across all fare families in the capacity calculation. This will allow reduction of the state variable to one dimension. Since (10) has the same form as (**??**), we can apply the MRT directly to get the one-dimensional fare family model.

$$\begin{aligned} TB_t(\mathbf{x}) = & \sum_{i=1}^{m} MD_{i,t} \cdot \max \left\{ f_i - UC_{i,t} - FM_{i,t} - BP_{t-1}(\mathbf{x}), 0 \right\} + q_t \mathbf{x} \cdot TB_{t-1}(\mathbf{x} - 1) + \\ & (1 - q_t \mathbf{x}) \, TB_{t-1}(\mathbf{x}) \end{aligned} \tag{11}$$

5 Conclusion

In this paper we have solved the joint overbooking and seat allocation problem for fare families by applying two transformations in succession. First the ECT, which transforms the refund cost from time of cancellation to time of booking. This reduces dimensionality from m dimension to n dimension (number of fare families). Subsequently the MRT, which transforms the dependent demand DP model to an equivalent independent demand DP model. This latter transformation provides an approximate 1-dimensional DP model, which is readily implemented in an RMS. The current model overcomes several of the inadequates of current overbooking models, such as dependence of Demand level, effects on class mix, effects of class specific refund costs, and class specific cancellation rates. We have performed simulation studies, that are not part of this technical report. The results will be published later as part of a journal paper. However initial results document significant revenue gain of 1%-3% compared to current industry standard.

Acknowledgements

We would like to thank Sara Skytte Olsen for her invaluable contributions to this research.

References

Jens Alstrup and Sren Boas. Bookingpolitik for fly med to passagertyper. Master's Thesis at the Technical University of Denmark, 1984.

Peter P. Belobaba. Application of a probabilistic decision model to airline seat inventory control. *Operations Research*, 37(2), March-April 1989.

S. E. Bodily and P.E. Pfeifer. Overbooking decision rules. *Omega*, 20:129 – 133, 1992.

Richard E. Chatwin. Multi-period airline overbooking with multiple fare classes. *Naval Research Logistics*, 43, 1996.

William L. Cooper, Tito Homem de Mello, and Anton J. Kleywegt. Models of the spiral-down effect in revenue management. *Operations Research*, 54(5), September-October 2006.

R. E. Curry. Optimal airline seat allocations with fare classes nested by origins and destinations. *Transportation Science*, 24(3):193 – 204, 1990.

Thomas Fiig, Karl Isler, Craig Hopperstad, and Peter P. Belobaba. Optimization of mixed fare structure: Theory and applications. *Journal of Revenue and Pricing Management*, 9, 2010.

Thomas Fiig, Nicolas Bondoux, Robert Hjorth, and Jesper Larsen. Joint overbooking and seat allocation for fare families. Presentation at AGIFORS Revenue Management Study Group Meeting, Frankfurt, Germany, 2016.

R. Klophaus and S. Pölt. Airline overbooking with dynamic spoilage costs. *Journal of Revenue and Pricing Management*, 6(1):9 – 18, 2006.

C. J. Lautenbacher and S. J. Stidham. The underlying markov decision process in the single-leg airline yield management problem. *Transportation Science*, 33:136 – 146, 1999.

Tak C. Lee and Marvin Hersh. A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Science*, 27(3), August 1993.

Ken Littlewood. Forecasting and control of passenger bookings. *Journal of Revenue and Pricing Management*, 4(2), 1972. Presentation given at the 12th AGIFORS Symposium in October 1972 - published in 2005.

R. M. Ratliff. Ideas on overbooking. Presentation at AGIFORS Reservation and Yield Management Study Group Meeting, Melbourne, Australia, 1998.

M. Rothstein and A. W. Stone. Passenger booking levels. In Proceedings 7th AGIFORS Symposium, American Airlines, New York., 1967.

Marvin Rothstein. An airline overbooking model. Technical report, University of Connecticut, June 1971.

Marvin Rothstein. OR and the airline overbooking problem. *Operations Research*, 33(2), March-April 1985.

E. Shlifer and Y. Vardi. An airline overbooking policy. *Transportation Science*, 9(2):101 – 114, 1975.

Janakiram Subramanian, Shaler Stidham Jr., and Conrad J. Lautenbacher. Airline yield management with overbooking, cancellations, and no-shows. *Transportation Science*, 33(2), 1999.

K. Talluri and G. van Ryzin. *The theory and practice of revenue management*. Springer Science & Business Media, 2006.

Kalyan T. Talluri and Garrett J. van Ryzin. Revenue management under a general discrete choice model of consumer behaviour. *Management Science*, 50(1), 2004.